Generating Temporal Data from User-Created Content to Detect Social Network Users with Depression

Akkapon Wongkoblap[1], MSc, Miguel A. Vadillo[2,3], PhD, Vasa Curcin[1,3], PhD
[1]Department of Informatics, King's College London, London, UK;
[2]Departamento de Psicología Básica, Universidad Autónoma de Madrid, Madrid, Spain;
[3]School of Population Health and Environmental Sciences, King's College London, London, UK

Abstract

Mental health issues are widely accepted as one of the most prominent health challenges in the world, with over 300 million people currently suffering from depression alone. With massive volumes of user-generated data on social networking platforms, researchers are increasingly using machine learning to determine whether this content can be used to detect mental health problems in users. This study aims to develop a deep learning model to classify users with depression via multiple instance learning, which can learn from user-level labels to identify post-level labels. By combining every possibility of posts label category, it can generate temporal posting profiles which can then be used to classify users with depression. This paper shows that there are clear differences in posting patterns between users with depression and non-depression, which is represented through the combined likelihood of posts label category.

1 Introduction

According to World Health Organization (WHO), the total number of people with depression globally was estimated to exceed 300 million in 2015[1]. Only in the United Kingdom, 16% of residents experience depression at some point in their lives[2]. Mental health problems are also predicted to cost 16.3 trillion USD between 2011 and 2030, through services, treatments, and a decline in productivity at work[3].

Diagnosis of mental disorder is a challenging task which can only be done by health professionals. For their disorder to be correctly ascertained, patients need to recall how they felt and what happened to them in the previous time period, which helps the clinicians obtain comprehensive background information. However, this method is time-consuming and error-prone as sometimes patients may not be able to correctly recall their experiences. An alternative way of collecting data on symptoms of mental illness is using a self-report questionnaire, whereby people can use the questions to obtain the data themselves, either as a one-off task or regularly.

With massive volumes of user-generated data being produced on social networking platforms, researchers are increasingly studying the how this content relates to users' mental health[4,5]. By focusing on user-generated messages, it is possible to screen for users with depression through social network data[6–8]. This alternative method of detecting depressed users from their published content can supplement the traditional diagnostics based on recollection data, by offering deeper insights into user's past activities, behaviours, and feelings.

Studies developing models to detect users with depression have typically used classical machine learning techniques e.g., support vector machines, regression, and random forests, with manual feature extraction and selection[6] and time-consuming data preparation. Deep learning techniques have been shown to successfully perform in a number of domains, but so far, there have been relatively few studies using deep neural networks for this task[9,10].

Individual posts published on social networks, related to users' activities, health, and feelings, are not annotated with labels and manual labelling of individual posts is a time-consuming task. Text classification models have been developed to label created content with categories[11], with some labelling posts related to user's mental health[4,9].

However, these labeling techniques have never been applied to classification models for predicting depression in social network users. This brings in a need for a predictive model which can automatically label every post with its categories and instantly detect users with depression from the labelled post patterns. In this paper, we make use of

multiple instance learning as it has been shown to successfully produce predictive models to classify sentiments on online review posts and identify sentiments on the sentence level by using only review-level labels[12,13].

This paper aims to investigate whether generated content from social networking users can provide changing patterns of content categories during observation periods. This raises two research questions:

(1) Can user-level labels transfer sentiment information to their unlabelled posts?

(2) Are there differences in posting patterns between users with depression and non-depression?

The main contributions of this paper are as follows:

(1) A model able to generate temporal data from user-generated text over observation time;

(2) A deep learning model to detect depression in social network users and label every post with sentiment information;

(3) Illustrating differences in the temporal data produced from social network posts between depressed and non-depressed users.

2 Methods

This section describes the dataset used in this study, the method to measure symptoms of depression and label our training set, the architecture of our model, and the experimental setup.

2.1 Dataset

This study used social network data from Facebook users to build a predictive model for detecting depression symptoms. The dataset was taken from the myPersonality project[14], obtained from participants who took a series of psychometric questionnaires, including the Center for Epidemiological Studies Depression (CES-D) form, and gave consent for this data to be shared. Some of them also gave permission for their Facebook profile data to be included. Their published content was downloaded and included in this dataset. This dataset was collected from 2007 until 2012.

The dataset contained 6,561 submissions of CES-D from 5,947 unique participants. Removing participants who withheld permission for their Facebook profiles to be included, 939 users remained in our dataset. To ensure that there are enough patterns to distinguish users between the two groups, users who published fewer than 100 posts on their timelines were excluded, leaving the total of 509 users in the final dataset.

2.2 Depression Symptom Measure

The CES-D questionnaire is one of standardised and popular tools to measure depressive symptoms of respondents who take it. It comprises 20 multiple answer questions, each of them asking respondents to rate how often they experienced certain symptoms over the past week, e.g. *"I was bothered by things that usually don't bother me"; "I felt I was just as good as other people"; "My sleep was restless"; "I enjoyed life"*. After every four items the wording of questions is reversed between positive and negative phrasings. Each answer has a score between 0 and 3 e.g., 0 = Rarely or none of the time (less than 1 day), 1 = Some or a little of the time (1-2 days), 2 = Occasionally or a moderate amount of time (3-4 days), and 3 = Most or all of the time (5-7 days). The total scores can then range between 0 and 60. Respondents with scores above the cut-off (typically between 16 and 24) are then classified as depressed. This study used the cut-off score at 22. This resulted in receiving 163 users without depression and 346 users with depression.

2.3 Predictive Model

The predictive model was completely trained using multiple instance learning (MIL) neural networks without manual feature engineering. The basic idea of MIL is to learn from a set of labelled bags, so the training does not require the individual labels in the training set instances, but only labelled bags of the training set, which sets it apart from supervised learning techniques that need to know the labels of all instances[15]. The MIL paradigm is suitable for our dataset, since it only had labels for the users but not for their individual posts.

The proposed architecture of our model is inspired by and follows the hierarchical attention network (HAN) introduced by Yang[16] and the multiple instance learning network (MILNET) proposed by Angelidis[13] and Kotzias[12]. The models have been shown to successfully perform sentiment analysis of online reviews. The concept of MILNET and its application can thus also be useful for developing a depression classifier. MILNET learns to analyse sentiment in a document from its encoding sentences or segments and then represents those as a document vector. Additionally, the model can identify the sentiment polarity of each segment of a given document. We adapt the MILNET approach by replacing segments with posted messages and a document vector with a user representation. Our proposed architecture consists of post encoder, post classification, user encoder, attention mechanism, and user classification (see Figure 1).

Post encoder

The first layer of our model transforms raw user post data into machine readable form. Word embedding was used to transform posts to word embedding metrics. A user publishes a number of posted messages $m_n = m_1 \cdots m_n$ and each post contains a sequence of words transformed to word embedding vector $We_i = We_1 \cdots We_i$. From the definition, x_p means the embedding metric in *p*-th post. The layer embeds words of posts to vectors:

$$x_n = m_n We_i.$$

After receiving word embedding vectors, convolutional neural networks (CNNs) are used to encode the vectors:

$$V_n = CNN(x_n).$$

Passing through the CNNs results in post representation V_n. The post encoding is then sent to the post classification part to perform sentiment analysis.

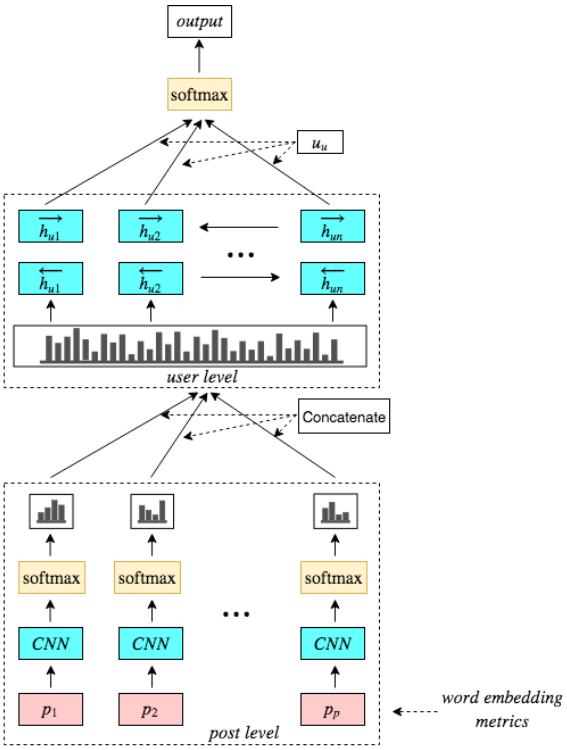

Figure 1. The architecture of multiple instance learning model for detecting users with depression

Post classification

After obtaining post representation, each post is classified based on whether it is mental health-related or related to another topic. To perform the classification, a softmax function[17]:

$$p_n = softmax(W_c V_n + b_c)$$

is applied to make separate predictions for every user post. The function generates post classification $p_n = p_1^C \cdots p_n^C$, where $C \in [0, 1]$ represents the sentiments with 1 denoting a mental health related post and 0 representing a non-mental-health related topic. The parameters W_c and b_c are learned and updated during the training step. After identifying individual post sentiments, every identified post label can be concatenated to generate a series of possibilities of post type.

User encoder

The series of post label predictions, called "user representation" in this study, is encoded to summarise the changing patterns of text categories over observation time. The user representation is received by combining all post label possibilities of a user. A bidirectional GRU is applied through the forward hidden state and the backward hidden state:

$$\overrightarrow{h_u} = \overrightarrow{GRU}(p_n)$$

$$\overleftarrow{h_u} = \overleftarrow{GRU}(p_n).$$

It produces vectors $\overrightarrow{h_u}$ and $\overleftarrow{h_u}$, which are then concatenated to $h_u = [\overrightarrow{h_u}, \overleftarrow{h_u}]$.

Attention mechanism

However, not all posts of a user convey a user characteristic. Some posts may contain cues that can be relevant to depression while others may not. For that purpose, we require the attention mechanism:

$$u_u = \tanh(W_w h_u + b_w)$$

$$\alpha_u = \frac{\exp(u_u^\top u_a)}{\sum_t \exp(u_u^\top u_a)}.$$

To be applied to reward posts that correctly represent the characteristic and are important to correctly detect a user with depression. The importance of a post is measured as the similarity of u_u with the context vector of post level u_a, which is learned and updated during the training step.

User classification

Finally, a user vector can be achieved through summarising all the information of post label possibilities of a user. The user vector v is computed as follows:

$$v = \sum_t \alpha_u p_n,$$

where α_u denotes the importance weight of a post and p_n represents the prediction of a type of the post. This results in obtaining a classifier to detect users with depression.

2.4 Experimental Setup

The proposed model was trained using the Keras library, the Python library for neural network APIs, with Tensorflow backend[18,19]. The word embedding dimensionality was set to 100. The wording embedding metrics were weighted and received from pre-trained word vectors of Glove[20]. The CNNs were applied with different window filter sizes. Adaptive moment estimation or Adam[21] was leveraged to train the model and categorical cross-entropy was applied to minimise loss. Every post of users in our dataset was tokenised, and the post length was limited to 100 tokens. The number of posts from every user was set to 500 to train the model. Users with fewer than 500 posts were

padded user metrics with the same length, filled with 0 values. Finally, the model was trained with cross-validation to build and test the model. Due to the highly imbalance dataset between users with and without depression, class weights were computed to weight the model during the training state.

3 Results

To report the performance of the proposed model, N-fold cross validation is used, splitting the dataset into n-equal small subsets using *n-1* subsets as training set and *one* subset as the test set. This is then iterated n times using each subset as the test set. Table 1 presents the results of accuracy, area under curve (AUC), precision, recall, and f1-score achieved by the model after training and testing with 5-fold cross validation. With the best performance, the model achieves the highest accuracy of 74.51%, while the average accuracy is 70.54%. The maximum results of precision, recall, and f1 score equal to 80%, 75%, and 73%, respectively. The model achieves the average results of precision of 68%, recall of 71%, and F1 score of 62%. It is noted that our dataset is highly imbalanced. Therefore, the baseline assessment would yield around accuracy of 68%., in the case of the model predicting only the majority class.

Table 1. Classification metrics of the baseline model

	Accuracy	**AUC**	**Precision**	**Recall**	**F1**
Fold 1	68.93%	67%	0.68	0.69	0.59
Fold 2	67.65%	64%	0.46	0.68	0.55
Fold 3	74.51%	77%	0.73	0.75	0.73
Fold 4	71.29%	71%	0.80	0.71	0.62
Fold 5	70.30%	73%	0.72	0.70	0.61
Average	**70.54%**	**70.40%**	**0.68**	**0.71**	**0.62**

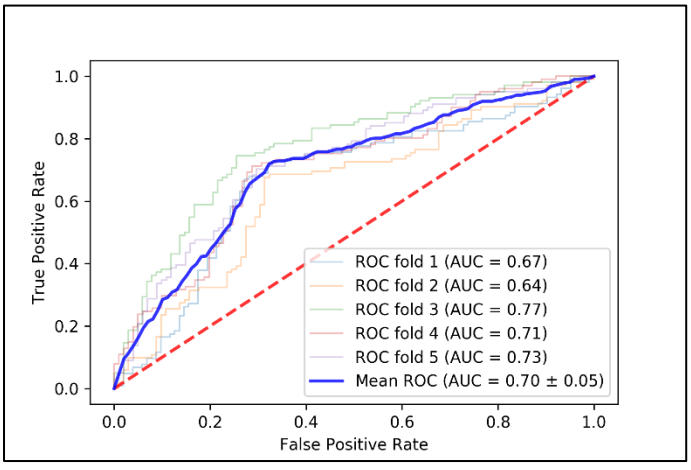

Figure 2. ROC curves of every testing fold of the baseline model

Figure 2 presents receiver operating characteristic (ROC) curves of evaluating the model with 5-fold cross validation. The model achieves average AUC of 70%, and the highest AUC is 77%. This highlights that our proposed model can perform better than chance. It can see that the results from the model present all the ROC curves above the red line or a random guess line.

4 Discussion

The purpose of this study is to develop a MIL predictive model to detect users with depression. The proposed model is evaluated and show performance results shown on the above. This finds out that the model can correctly label users with the maximum accuracy of 74.51%. The best model is extracted and visualised predicted weight values from its hidden layers. We extracted labelled values of all posts from post classification layer. The purpose of the layer is to label every given post of a user into either general or mental health related text. The possibility of being general or mental health related text are calculated and estimated from given labels of users.

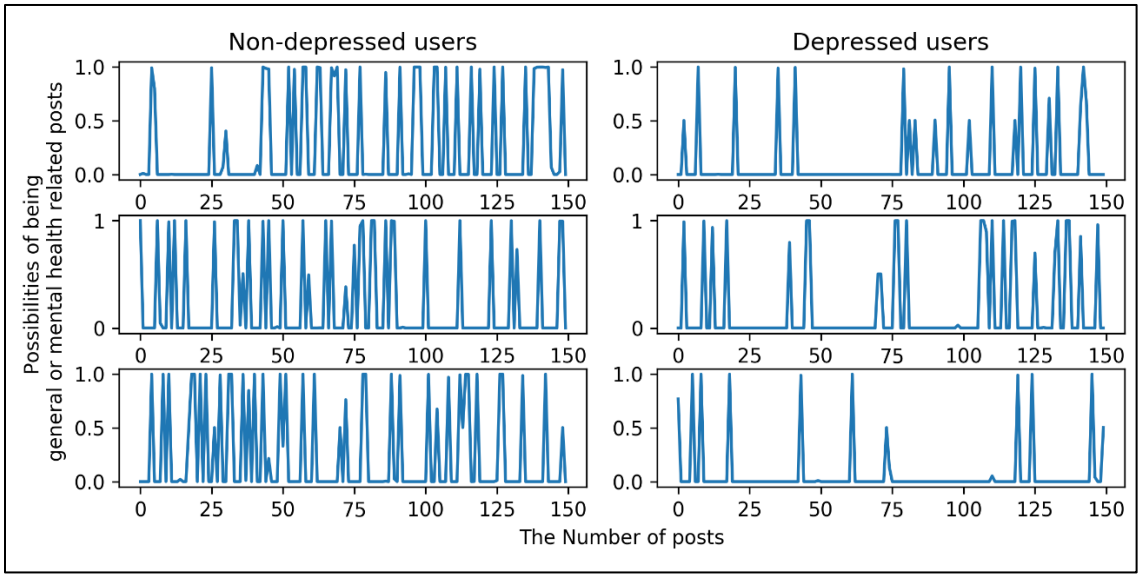

Figure 3. The patterns of general and mental-health related posts in 3 depressed and 3 non-depressed users. The *y*-axes denotes the likelihood of the post topic being mental health related. The *x*-axis shows the number of posts.

Figure 3 shows changing patterns of publishing posts related to general or mental health topics. The figure represents the patterns from 6 random users with depression and non-depression. The *y*-axis of the chart shows the possibilities of being another topic denoted with 0 and mental health related text denoted with 1, and the *x*-axis is the number of posts. These patterns are generated from users who were correctly labelled by our proposed model. On the left hand side, it can be seen that users with non-depression tended to have fluctuating changes over time. Considering users with depression, the changing patterns were more stable. This highlights that the model could use the patterns to distinguish between the two groups of users.

We further investigated whether post content of the users can be correctly labelled with mental health or general topics. We found that our model could not correctly label the posts. It is possible that our dataset was not sufficiently large to train both user level and post level classifiers. Another reason is that the model used long sequences of posts. We used 500 posts per a user to train our proposed model. In comparison, the original model from Angelidis study[13], used around 8 to 14 segments (equivalent to posts in our study) per single document, which is much shorter than our sequences. In Angelidis dataset, they used more data (~300k documents) to train their model, while our dataset had 509 users. They also used more labels e.g., document-level sentiment classes ranged from 1 to 10, while we used only non-depression and depression labels.

5 Conclusion

In this paper, we developed a MIL predictive model to detect social network users with depression. It found that our proposed model achieved the maximum accuracy of 74.51% and the highest precision of 80% in detecting depressed users from their social network created content with additionally generating changing patterns or user representation. This study highlights that the model could provide insights into the changes of generated content. To

our best knowledge, this study is the first one to apply the MIL model and generate temporal data from created text to detect social network users with depression.

Our proposed model learnt a user representation by transforming words into embedding vectors of posts to learn the importance of post representation and aggregate the importance to user representation. The user representation or temporal data of posting were then used to distinguish between the two classes.

The model could potentially be applied to structured and unstructured text and map it to temporal data, which can provide better understandings of changing patterns of text over observation time. The transformation of text data to temporal data may have a considerable impact to health care research, e.g., transforming health-related text or electronic health records (EHRs) to temporal data to provide patterns of patients to a doctor.

This study is not free from limitations. Our dataset was highly unbalanced, as it had a higher number of depressed users than what is found in general population. Another limitation is that the model was trained with a relatively few of users, and the performance may be improved if it is trained on a larger sample.

As further work, we plan to improve the proposed model via transfer learning. The idea is to train the post classification part with other labelled text to boost the model capacity to classify text related to mental health content. This may improve the performance of detecting users with depression and provide us with more insights into how generated text content changes over time.

References

1. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. http://apps.who.int/iris/bitstream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf. Published 2017. Accessed September 27, 2017.

2. Mental Health Foundation. *Fundamental Facts About Mental Health 2016*. London: Mental Health Foundation; 2016.

3. Bloom DE, Cafiero E, Jané-Llopis E, Abrahams-Gessel S, Bloom RL, Fathima S, Feigl BA, Gaziano T, Mowafi M, Pandya A, Prettner K, Rosenberg L, Seligman B, Stein AZ, Weinstein C. The Global Economic Burden of Noncommunicable Diseases. World Economic Forum. doi:6CSThUnbF

4. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJP, Dobson RJB, Dutta R. Characterisation of mental health conditions in social media using Informed Deep Learning. *Sci Rep*. 2017;7:45141. doi:10.1038/srep45141

5. Yazdavar AH, Al-Olimat HS, Ebrahimi M, Bajaj G, Banerjee T, Thirunarayan K, Pathak J, Sheth A. Semi-Supervised Approach to Monitoring Clinical Depressive Symptoms in Social Media. In: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASONAM '17*. New York, New York, USA: ACM Press; 2017:1191-1198. doi:10.1145/3110025.3123028

6. Wongkoblap A, Vadillo MA, Curcin V. Researching Mental Health Disorders in the Era of Social Media: Systematic Review. *J Med Internet Res*. 2017;19(6):e228. doi:10.2196/jmir.7215

7. De Choudhury M, Gamon M. Predicting Depression via Social Media. *Proc Seventh Int AAAI Conf Weblogs Soc Media*. 2013;2:128-137. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6124/6351.

8. Coppersmith G, Dredze M, Harman C. Quantifying Mental Health Signals in Twitter. *Proc Work Comput Linguist Clin Psychol From Linguist Signal to Clin Real*. 2014:51-60. doi:10.3115/v1/W14-3207

9. Ive J, Gkotsis G, Dutta R, Stewart R, Velupillai S. Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health. In: *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. Association for Computational Linguistics; 2018:69-77. http://aclweb.org/anthology/W18-0607.

10. Amir S, Coppersmith G, Carvalho P, Silva MJ, Wallace BC. Quantifying Mental Health from Social Media with Neural User Embeddings. In: Doshi-Velez F, Fackler J, Kale D, Ranganath R, Wallace B, Wiens J, eds. *Proceedings of the 2nd Machine Learning for Healthcare Conference*. Vol 68. Proceedings of Machine Learning Research. Boston, Massachusetts: PMLR; 2017:306-321. http://proceedings.mlr.press/v68/amir17a.html.

11. Wang P, Xu J, Xu B, Liu C, Zhang H, Wang F, Hao H. Semantic Clustering and Convolutional Neural Network for Short Text Categorization. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Beijing, China: Association for Computational Linguistics; 2015:352-357. http://www.aclweb.org/anthology/P15-2058.

12. Kotzias D, Denil M, de Freitas N, Smyth P. From Group to Individual Labels Using Deep Features. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*. New York, New York, USA: ACM Press; 2015:597-606. doi:10.1145/2783258.2783380

13. Angelidis S, Lapata M. Multiple Instance Learning Networks for Fine-Grained Sentiment Analysis. *Trans Assoc Comput Linguist*. 2018;6:17-31. https://www.transacl.org/ojs/index.php/tacl/article/view/1225.

14. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *Am Psychol*. 2015;70(6):543-556. doi:10.1037/a0039210

15. Carbonneau M-A, Cheplygina V, Granger E, Gagnon G. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognit*. 2018;77:329-353. doi:10.1016/j.patcog.2017.10.009

16. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ; 2016:1480-1489.

17. Bishop CM. *Pattern Recognition and Machine Learning*. Springer; 2006.

18. Chollet F, others. Keras. 2015.

19. Abadi M, Barham P, Chen J, et al. TensorFlow: A System for Large-scale Machine Learning. In: *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*. OSDI'16. Berkeley, CA, USA: USENIX Association; 2016:265-283. http://dl.acm.org/citation.cfm?id=3026877.3026899.

20. Pennington J, Socher R, Manning CD. GloVe: Global Vectors for Word Representation. In: *Empirical Methods in Natural Language Processing (EMNLP)*. ; 2014:1532-1543. http://www.aclweb.org/anthology/D14-1162.

21. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *CoRR*. 2014;abs/1412.6. http://arxiv.org/abs/1412.6980.